THE REVISED SIMPLEX METHOD

William Orchard-Hays

P-911

August 6, 1956

The RAND Corporation
1700 MAIN ST. • SANTA MONICA • CALIFORNIA

SUMMARY

A revision of the simplex method is presented which makes explicit use of columns of the restraint coefficients associated with a basic set of variables. The development is based on the single assumption of linearly independent restraint equations. An algebraic method of resolving degeneracy is given in conclusion.( )

CONTENTS

THE REVISED SIMPLEX METHOD

Wm. Orchard-Hays

INTRODUCTION

Although this course was designed for those interested in LP computations, it was felt that a considerable amount of theoretical background should be included. While this might be in the nature of a review for some people, I believe you will all agree that the material presented thus far has been very instructive and has provided us with a necessary solid and common foundation upon which to continue. In fact, a natural question now might be, that except for computational tricks, what is there to say further? Indeed we have enough mathematical theory at this point to carry out the computations required by a given LP model, at least with a little luck and provided it is not too large. There is, of course, a great deal that has not and can not be said in our limited time about model formulation. For example, linear approximations to convex functionals are interesting mathematically. The transportation problem, which will be discussed in the next two lectures, has at least one result of primary importance. But insofar as a general method of solution is concerned, the preceding theory is more or less complete except for the matter of degeneracy. Nevertheless, we have not yet begun, in reality, to speak about the practical computational difficulties that confront us nor how we propose to resolve them.

These matters will be taken up in more detail during the second three days of the course. In preparation for this, the present lecture takes up a revision of the simplex method which is computationally desirable.

In later lectures, we will find it convenient to use a slightly different notation than was used in the last one. Also, in the transportation problem, you will find an array of quantities with double subscripts used in a different sense than in the general simplex method. Hence this seems like a good place to introduce the notation to be used later on. To those of you familiar with matrix theory, the viewpoint taken in some of what follows may seem narrow but the presentation is not intended to be general from a matrix theoretical standpoint but only to suit our purposes assuming as little background as possible. Aitken, in his excellent little book "Determinants and Matrices," says: "It would be intolerably tedious if, whenever we had occasion to manipulate sets of equations or to refer to properties of the coefficients, we had to write either the equations or the scheme of coefficients in full. The need for an abbreviated notation was early felt..." He also refers to matrix notation as the shorthand of algebra "expressed at still shorter hand" and as a code. This is the viewpoint which will be useful here, rather than that of abstract algebra.

LINEAR EQUATIONS; NOTATION

A linear equation in n unknowns is one of the form

$$(1) \qquad a_1x_1 + a_2x_2 + \dots + a_nx_n = b$$

where the a's and b are given numbers. The subscripts are called indices and are used to distinguish different numbers used in the same or similar ways.

For reasons that will appear presently, let us put the indices for the unknowns as superscripts (since these are linear equations, there is no confusion with exponents)

$$(1') \qquad a_1x^1 + a_2x^2 + \dots + a_nx^n = b.$$

Suppose we have several such equations, say m of them, which are to be satisfied simultaneously, that is for the same set of x's. We can distinguish these by another index on the a's and b's which we can also write as a superscript:

$$(2) \qquad \begin{array}{c} a_1^1x^1 + a_2^1x^2 + \dots + a_n^1x^n = b^1 \\ a_1^2x^1 + a_2^2x^2 + \dots + a_n^2x^n = b^2 \\ \vdots \\ a_1^mx^1 + a_2^mx^2 + \dots + a_n^mx^n = b^m \quad . \end{array}$$

At this point it becomes clear that some sort of condensed but systematic notation is highly desirable if not actually necessary. Several schemes have been used but they are by no means standardized. Out of six standard texts on my shelf, there are six different notations for this purpose. However, we might, for example, write the a_j^i elements of (2) in detached form, i.e. without the x's and plus signs, but in an array which maintains

their relative positions, and designate this array by the single capital letter A. Similarly write the b's from the right side of (2) in a column and call this column B.

$$A = \begin{bmatrix} a_1^1 & a_2^1 & \cdots & a_n^1 \\ a_1^2 & a_2^2 & \cdots & a_n^2 \\ & & \vdots & \\ a_1^m & a_2^m & \cdots & a_n^m \end{bmatrix} \qquad B = \begin{bmatrix} b^1 \\ b^2 \\ \vdots \\ b^m \end{bmatrix}$$

The array A is called an mxn matrix. B is an mX1 matrix. A and B are now considered as abstract quantities, i.e. a new kind of numerical animal. There is a whole algebra based on these, in which multiplication is defined in a one-sided way. The left factor is always read row-wise and the right factor column-wise. That is, one multiplies a row on the left by a column on the right by summing the products of elements with the same numerical index, one from each array. The resulting single number is the element of the right-hand side with the same index, here a superscript, as the row. So if we write the matrix equation

$$(3) \qquad AX = B$$

X must be a column and hence has its elements distinguished by superscripts in the present system:

$$X = \begin{bmatrix} x^1 \\ x^2 \\ \vdots \\ x^n \end{bmatrix}$$

Note that X is an $n \times 1$ matrix. The single equation (3) now means exactly the same as the system (2). We have certainly achieved a condensation in notation. It tends, however, to go a little too far and obscure the mass of numbers with which we are really dealing. It also makes it hard to refer to a particular element, a particular column or a particular row without having four kinds of symbols.

The detached columns B and X are called vectors. Detached rows are also called vectors and they do not operate in the same way as columns. If the distinction between rows and columns is not maintained, chaos results.

The other common shorthand for systems of linear equations uses the summation sign, Σ, as follows.

$$(4) \qquad \sum_{j=1}^{n} a^i_j \, x^j = b^i \qquad (i = 1, \ldots, m)$$

which reads: The sum of the products $a^i_j \, x^j$ taken for fixed i and $j = 1, \ldots, n$ gives b^i and this is true for $i = 1, \ldots, m$. Equation (4) has the advantages of retaining individual symbols for each element and of distinguishing clearly rows and columns. We will therefore dispense with the capital letters of the abstract notation (3) and simply write a^i_j, x^j, b^i to mean the arrays called above A, X, B respectively. Similarly, a single

row of elements c_1, c_2, ..., c_n will be designated simply c_i.

Note that the letters used for indices are unimportant. We could just as well write a_k^h or even a_i^j in place of a_j^i. But we sometimes want to refer to a particular row, column or element in a matrix. We will reserve the two letters r and s for this purpose. Thus a_j^r means the r^{th} row of a_j^i, a_s^i means the s^{th} column of a_j^i and a_s^r means the element of a_j^i in row r and column s. In other words, when r and s are used, we will understand that they are limited to one value and not the whole range 1, ..., m or 1, ..., n. In this lecture we will occasionally use q and t in this way also.

CENTRAL MATHEMATICAL PROBLEM OF LP

An LP problem now takes the form

Given: a_j^o , a_j^i , b^i $(i = 1, \ldots, m; \; j = 1, \ldots, n)$

Find: $x^j \geq 0$ such that

$$\sum_{j=1}^{n} a_j^o \, x^j \text{ is minimum and}$$

$$\sum_{j=1}^{n} a_j^i \, x^j = b^i \quad (i = 1, \ldots, m; \; m < n).$$

It seems worthwhile to point out here what the nature of the problem really is. We are seeking a non-negative solution to a system of linear equations which minimizes a linear form. The important fact is that there exists a finite number of feasible solutions in terms of which all others can be expressed. These are precisely the basic feasible solutions. The number

of these, however, may be very large, not exceeding the number of combinations of n things taken m at a time,

$$\binom{n}{m} = \frac{n!}{m!\,(n-m)!} \,.$$

Even if only a small fraction of these is feasible, there may be an awful lot of them. For example, if n = 100, m = 50 and only about 1 in a billion combinations is feasible, this is still about 10^{20} basic solutions. The fact that the simplex method has never been known to iterate through a significant fraction of this number of solutions to reach one that is optimal is one of the more pleasant phenomena of life.

LINEAR INDEPENDENCE

Suppose it is possible to find elements of a row vector λ_i $(i = 1, \ldots, m)$ not all zero, such that $\sum_{i=1}^{m} \lambda_i \, a^i_j = 0$ for all j. Then the rows of the matrix a^i_j are said to be linearly dependent, that is some one of them can be represented as a linear combination of the others. If this cannot be done, then the rows are said to be linearly independent. For the moment, let us assume that the rows are independent since, in practice, there is a way of insuring this. It is the first requirement for obtaining a solution, since otherwise one does not necessarily exist for an arbitrary right-hand side, b^i. The significance of linearly independent rows is that there is always a solution for any given b^i, though not necessarily with non-negative x^j. To see that no solution exists for certain b^i if the rows are dependent, find some set of λ_i not all zero such that

$\sum_i \lambda_i a^i_j = 0$ for all j. Next set $b^i = \lambda_i$ $(i = 1, \ldots, m)$. Now adding together all equations multiplied by λ_i gives

$$0 = \sum_j \sum_i \lambda_i a^i_j x^j = \sum_i \lambda_i \lambda^i > 0$$

which is obviously impossible for any x^j. The last inequality holds because the right-hand side is a sum of squares not all zero.

The way in which linear independence comes into play in our problem is that the independence of the m rows implies that there is at least one set of m columns which are also linearly independent and hence form a basis, that is a set of columns in terms of which any right-hand side can be represented. This is usually proved in terms of determinants or linear transformations. We will prove it by elimination, which is closely allied to the method of determinants (in fact, they used to be called eliminants) but which is more apropos of the simplex method.

<u>ELEMENTARY ROW TRANSFORMATIONS</u>

<u>Definition</u>: An elementary row transformation (<u>E.R.T.</u>) on an mxn matrix a^i_j is one which replaces some row $i = r$ with

$$\bar{a}^r_j = \lambda a^r_j + \mu a^s_j \quad (\lambda \neq 0,\ s \neq r).$$

<u>Theorem 1</u>: E.R.T.s are reversible and the inverse transformation is an E.R.T.

Proof: Clearly if $\bar{a}^r_j = \lambda a^r_j + \mu a^s_j$ $(\lambda \neq 0,\ s \neq r)$ then

$$a^r_j = \frac{1}{\lambda}\bar{a}^r_j - \frac{\mu}{\lambda} a^s_j \quad \text{uniquely}$$

and the latter is also an E.R.T.

<u>Theorem 2</u>: E.R.T.s preserve linear independence of the rows.

Proof: Suppose $\sum_i \lambda_i a^i_j = 0$ for all j implies $\lambda_i = 0$ for all i. Replace a^r_j with $\bar{a}^r_j = \lambda a^r_j + \mu a^s_j$ $(\lambda \neq 0,\ s \neq r)$. Now let μ_i be such that

$$\sum_{i \neq r} \mu_i a^i_j + \mu_r \bar{a}^r_j = 0 \quad \text{for all } j.$$

Then

$$\sum_{\substack{i \neq r \\ i \neq s}} \mu_i a^i_j + (\mu_s + \mu\mu_r)\, a^s_j + (\mu_r \lambda)\, a^r_j = 0$$

which, by above assumption of independence, implies

$$\mu_i = 0 \quad (i \neq r, s)$$

$$\mu_s + \mu\mu_r = 0 \quad \text{and} \quad \mu_r \lambda = 0.$$

Since $\lambda \neq 0$, $\mu_r = 0$ and hence $\mu_s = 0$. also. Thus the transformed system still has independent rows.

<u>Definition</u>: A <u>unit column vector</u> δ^i_r for $1 \leq r \leq m$ is defined by

$$\delta^i_r = \begin{cases} 0 & \text{if } i \neq r \\ 1 & \text{if } i = r \end{cases}.$$

The set of unit column vectors for r = 1, 2, ..., m, in that order, form the <u>identity matrix</u> of order m, δ^i_h $(i, h = 1, \ldots, m)$.

NOTE: δ^i_h is usually called the Kronecker delta. In our notation, it is the identity matrix also.

<u>Theorem 3</u>: Any column in a^i_j with at least one non-zero element can be transformed into a unit column vector by E.R.T.s.

Proof: Suppose that for column s, $a_s^r \neq 0$. Then row r can be replaced with $\bar{a}_j^r = \frac{1}{a_s^r} a_j^r$ $(j = 1, \ldots, n)$. Obviously $\bar{a}_s^r = 1$. Then every row $i \neq r$ can be replaced with

$$\bar{a}_j^i = a_j^i - a_s^i \bar{a}_j^r \qquad (i \neq r).$$

Now clearly for $j = s$, $\bar{a}_s^i = a_s^i - a_s^i = 0$ $(i \neq r)$, or, for all i, $\bar{a}_s^i = \delta_r^i$ as was to be shown.

The element a_s^r in the above proof is called the pivot element.

Corollary: If any δ_q^i $(q \neq r)$ appeared in a_j^i, then it remains intact in $\bar{a}_j^i$.

Proof:
Suppose $a_t^i = \delta_q^i$. Then $t \neq s$ if $q \neq r$. Hence

$$\bar{a}_t^r = \frac{1}{a_s^r} a_t^r = 0$$

$$\bar{a}_t^i = a_t^i - a_s^i \bar{a}_t^r = a_t^i - a_s^i \cdot 0 = a_t^i \qquad (i \neq r) \quad .$$

Theorem 4: If the rows of a_j^i are linearly independent then the matrix can be transformed into b_j^i by E.R.T.s so that every column of δ_h^i appears somewhere in b_j^i.

Proof: Choose some $a_s^r \neq 0$. By Th. 3, a_j^i can be transformed by E.R.T.s into $\bar{a}_j^i$ where $\bar{a}_s^i = \delta_r^i$. Now no row of $\bar{a}_j^i$ can contain all zeros. For, suppose $\bar{a}_j^q = 0$ for all j. Then setting $\lambda_q = 1$, $\lambda_i = 0$ for $i \neq q$,

$$\sum_{i=1}^{m} \lambda_i \bar{a}_j^i = 0 \text{ for all } j$$

and the rows are linearly dependent. Hence, by Th. 2,

the rows of a^i_j were dependent, contrary to assumption. Hence we can choose some $q \neq r$, $t \neq s$ for which $\bar{a}^q_t \neq 0$ and transform $\bar{a}^i_j$ by E.R.T.s into $\bar{\bar{a}}^i_j$ with $\bar{\bar{a}}^i_t = \delta^i_q$. However, any column of δ^i_h $(h \neq q)$ which appeared in $\bar{a}^i_j$ will remain undisturbed in $\bar{\bar{a}}^i_j$ by the corollary to Th. 3. This process can clearly be repeated until every row has been chosen, each with a different column, so that we arrive at the matrix b^i_j as required.

Corollary: If $m > n$, then the rows of a^i_j are dependent.

Proof: Even assuming we can choose n rows, at that point all elements of the transformed matrix will be zero except the pivot elements already chosen. Consequently the rows of the transformed matrix are dependent, hence so are those of a^i_j.

We need a few facts which are practically obvious. We will state them as lemmas.

Lemma 1: Both the rows and the columns of δ^i_h are linearly independent.

Lemma 2: If the same E.R.T.s are applied to the right hand side of a system of linear equations as are applied to the matrix of coefficients, each solution of the original system is a solution of the transformed system.

Proof: Suppose $\sum_j a^i_j x^j = b^i$ $(i = 1, \ldots, m)$ is satisfied for some set of values x^j. Then $\sum_j (\lambda a^r_j + \mu a^s_j) x^j = \lambda b^r + \mu b^s$ and thus the r^{th} equation of the system

obtained by applying an E.R.T. is satisfied. Clearly the others are also. Applying this argument repeatedly proves the lemma.

Lemma 3: $\sum_j \delta^i_j b^j = b^i$;

$\sum_i c_i \delta^i_j = c_j$;

$\sum_k \delta^i_k \delta^k_j = \delta^i_j$.

These all follow immediately from the definition of δ^i_h.

Definition: An m×m matrix consisting of linearly independent rows is said to be non-singular.

Theorem 5: If an m×m matrix μ^i_h has linearly independent rows, it also has linearly independent columns.

Proof: Let β^h be such that $\sum_h \mu^i_h \beta^h = 0$ for all i. We can transform μ^i_h by E.R.T.s to π^i_h which differs from δ^i_h only in the order of the columns. Applying the same E.R.T.s to the right hand side maintains the equalities without changing the β^h. But since E.R.T.s applied to all zeros obviously give all zeros, the β^h must all be zero. Hence the columns of μ^i_h are linearly independent.

Theorem 6: If the m×m matrix μ^i_h is non-singular, then the system of equations

$$\sum_{h=1}^{m} \mu^i_h x^h = b^i$$

has a unique solution $x^h = \beta^h$ for any given b^i whatsoever.

Proof: By E.R.T.s, μ^i_h can be transformed to $\overline{\mu}^i_h$ which differs from δ^i_h at most in the order of the columns. At the same time, b^i can be transformed to $\overline{b}^i$. If $\overline{\mu}^i_s = \delta^i_r$, then the value of x^s must be $\overline{b}^r$ and this is true for each $s = 1, \ldots, m$. Furthermore, for each such s, a different value of r will be chosen so that $r = \phi(s)$ is a permutation of the indices $s = 1, \ldots, m$. Hence the solution of the transformed system is $x^h = \overline{b}^{\phi(h)}$, uniquely. But we can retrace our steps to the original system by the inverse E.R.T.s applied in reverse order without changing the values of the x^h. Hence $x^h = \beta^h = \overline{b}^{\phi(h)}$ is a solution; it is unique by Lemma 2.

Corollary to Ths. 5 and 6: If the mxm matrix μ^i_h is non-singular, then the system of equations

$$\sum_{i=1}^{m} y_i \mu^i_h = c_h$$

has a unique solution $y_i = \gamma_h$ for any given c_h whatsoever.

Proof: The whole system can be transposed to the form of Th. 6. That is, let $u^h_i = \mu^i_h$ and replace y_i with y^i, c_h with c^h. Then we have

$$\sum_{i=1}^{m} u^h_i y^i = c^h.$$

By Th. 5, the rows of u^h_i are independent.

We now observe that we can solve a system simultaneously for several right-hand sides. Suppose we wish to find a solution to

$$\sum_{j=1}^{n} a_j^i \, x_k^j = b_k^i \qquad (i = 1, \ldots, m)$$

for p different right-hand sides $b_1^i, b_2^i, \ldots, b_p^i$. Then we can write x_k^j as an nxp matrix and b_k^i as an mxp matrix. Each column k of x_k^j will be the solution to the same column k of b_k^i. Hence we see that, in general, an m n matrix times an nxp matrix gives an mxp matrix. We will now state the general rule of

<u>Multiplication of Matrices</u>: If a_j^i is an mXn matrix and b_j^i is an nxp matrix then the mxp matrix c_j^i given by the rule of multiplication

$$\sum_{k=1}^{n} a_k^i \, b_j^k = c_j^i$$

is called the product of a_j^i and b_j^i in that order.

<u>Lemma 4</u>: Multiplication of matrices is associative, that is, if a_j^i is ℓxm, b_j^i is mXn and c_j^i is nXp, then

$$\sum_{h=1}^{m} a_h^i \left[\sum_{k=1}^{n} b_k^h \, c_j^k \right] = \sum_{k=1}^{n} \left[\sum_{h=1}^{m} a_h^i \, b_k^h \right] c_j^k \, .$$

We leave this for the reader to prove to his own satisfaction. It is a well known fact that multiplication of matrices is not commutative, that is, even if a_j^i is mxn and b_j^i is nxm (why is this necessary?) that

$$\sum_k a^i_k \, b^k_j \quad \text{is not in general equal to} \quad \sum_k b^i_k \, a^k_j \, .$$

Obviously, if $m \neq n$ then the two products are not even of the same order. But even when $m = n$, the two are not generally equal.

Theorem 7: If the $m \times m$ matrix μ^i_h is non-singular, then there exists a unique $m \times m$ matrix π^i_h such that

$$(5) \quad \sum_k \mu^i_k \, \pi^k_h = \delta^i_h \, .$$

Proof: Using, for $t = 1, \ldots, m$, the t^{th} column of δ^i_h as the right hand side, we can obtain a unique solution π^k_t by Th. 6.

The matrix π^i_h is called the <u>inverse</u> of μ^i_h .

<u>Theorem 8</u>: A non-singular matrix commutes with its inverse.

Proof: Multiply both sides of (5) on the left by π^j_i .

By lemma 4, we can combine summation signs:

$$\sum_{i,k} \pi^j_i \, \mu^i_k \, \pi^k_h = \sum_i \pi^j_i \, \delta^i_h = \pi^j_h$$

Let $\sum_i \pi^j_i \mu^i_k = d^j_k$.

Then $\sum_k d^j_k \, \pi^k_h = \pi^j_h$

Hence, by the Corollary to Th. 5 and 6, $d^j_k = \delta^j_k$ which with (5) proves the statement.

THE BASIS IN THE SIMPLEX METHOD

Suppose that in solving an LP problem in which the rows are linearly independent we have arrived at a basic solution in which a certain set B of m indices j specifies the basic variables. The columns a_j^i ($j \in B$) are referred to as a basis. It is clear that any matrix formed from these columns is non-singular. We do not know, in general, in what order these columns were chosen nor what columns of δ_h^i they become in canonical form, i.e. after application of the E.R.T.s. Indeed the set of E.R.T.s necessary to get us to this point is not at all unique. Of course, if we have a machine code, which must follow certain detailed rules, and we feed it the same problem, ordered in the same way, twice, then it will, we hope, do the same identical operations both times. But a slight change in the rules or the order of the columns will lead to the same result along a different path. We can avoid a good deal of confusion by distinguishing between the index which is the name of a variable and the index which denotes the place its coefficient column occupies in a particular ordering of the basis.

Let us pretend that we have lost all information about our basic solution except the list of name indices in B. Since the basic solution is unique, any way in which we reconstruct it is valid, that is, an interchange of two columns merely interchanges the two corresponding values in the final solution column. Let us denote this column by β^h. Hence, which x^s any particular β^r is the value of, depends on what column a_s^i ($s \in B$) becomes δ_r^i

when we transform the basis. Let us denote this correspondence by $s = \phi(r)$. Thus $\phi(h)$ is a list of m name indices ordered on $h = 1, \ldots, m$. We will refer to this list $\phi(h)$ as the <u>basis headings</u>. We do not think of it either as a column or a row but merely a list denoting a correspondence.

Let us now attempt to reconstruct our lost solution from the original data and the set B of indices j. Let us form an mx2m matrix in which the first m columns are the a_j^i for j in B and the last m are δ_h^i. We will associate the $\phi(h)$ with the columns of δ_h^i and make $\phi(h) = 0$ initially.

From the first m columns, choose some $a_s^r \neq 0$. By E.R.T.s on the whole array we can transform a_s^i into δ_r^i. At the same time we apply the same E.R.T.s to the right-hand side b^i. We also change $\phi(r)$ to s. We can now choose another pivot element from the first m columns, say $\bar{a}_{s'}^{r'} \neq 0$, and transform column s' to $\delta_{r'}^i$, $(r' \neq r,\ s' \neq s)$. We make $\phi(r') = s'$. Proceeding in this manner, we can change every one of the first m columns into unit column vectors and determine the complete list of basis headings.

Now let us suppose the first m columns had been ordered in the way the list $j = \phi(h)$ finally turned out. Then these columns would have been transformed into the unit columns in proper order, i.e. δ_h^i. The last m columns, initially δ_h^i, have been transformed into the solution to

$$(6) \qquad \sum_{h=1}^{m} a_{\phi(h)}^i \, \pi_j^h = \delta_j^i \quad (i,j = 1, \ldots, m)$$

as is seen in the proof of Th. 6. (The ϕ used there was the inverse of the ϕ used here.) At the same time, the column b^i has been transformed into the solution to

$$(7) \qquad \sum_{h=1}^{m} a^i_{\phi(h)} \beta^h = b^i \qquad (i = 1, \ldots, m)$$

Hence the last m columns have become the inverse of the ordered basis $a^i_{\phi(h)}$ $(h = 1, \ldots, m)$ and the value of $x^{\phi(h)} = \beta^h$.

THE TRANSFORMATION OF THE COST ROW

We have neglected thus far the coefficients a^o_j of the objective form. However, in reducing the basis to canonical form, there is nothing to prevent us from performing extra E.R.T.s on this row to make each coefficient associated with the basis vanish. The total effect of all these E.R.T.s is equivalent to solving for variables π_i in a transposed system

$$(8) \qquad a^o_{\phi(h)} + \sum_{i=1}^{m} \pi_i \, a^i_{\phi(h)} = 0 \qquad (h = 1, \ldots, m).$$

Putting $a^o_{\phi(h)}$ on the right in (8) and multiplying both sides by π^h_j on the right gives

$$\sum_{i,h} \pi_i \, a^i_{\phi(h)} \, \pi^h_j = - \sum_h a^o_{\phi(h)} \, \pi^h_j$$

or by (6)

$$\sum_i \pi_i \, \delta^i_j = \pi_j = - \sum_h a^o_{\phi(h)} \, \pi^h_j \, .$$

Hence, the transformed cost coefficients are given by

$$(9) \qquad d_j = a^o_j + \sum_{i=1}^{m} \pi_i \, a^i_j \qquad (j = 1, \ldots, n)$$

where $d_j = 0$ for j in B by definition. Note that the order of the π_i in (8) is not affected by the ordering $j = \phi(h)$ of the columns a_j^i.

The value of the basic solution is

$$z = \sum_{h=1}^{m} a_{\phi(h)}^{o} \beta^{h} = \sum_{h,i} a_{\phi(h)}^{o} \pi_i^h b^i = - \sum_i \pi_i b^i .$$

THE REVISED SIMPLEX METHOD

The row vector π_i in (8) is called the <u>pricing vector</u>. As seen above, it can be generated at the same time as the inverse π_i^h is computed, in fact it is convenient to think of π_i as π_i^o. The revised simplex method consists of using the pricing vector to determine a vector a_s^i $(s \not\in B)$ to bring into the basis, and of using the inverse to represent a_s^i in terms of the basis. It is clear by this time, of course, that the transformed column $\bar{a}_s^i$ used in the original simplex method is nothing but the values of the unknowns α_s^i in the equations

$$\sum_{h=1}^{m} a_{\phi(h)}^{i} \alpha_s^h = a_s^i \qquad (i = 1, \ldots, m) .$$

Whence

$$\sum_{i,h} \pi_i^k a_{\phi(h)}^{i} \alpha_s^h = \alpha_s^k = \sum_i \pi_i^k a_s^i .$$

The index s is chosen as usual by the rule

$$d_s = \min_j d_j < 0$$

taking smallest index, say, in case of ties. If all $d_j \geq 0$, then the present solution is optimal.

The rule for choosing the index r of the vector to go out of the basis is

(10) $$\theta^r = \min_{\alpha_s^i > 0} \left\{ \frac{\beta^i}{\alpha_s^i} \right\} \text{ if any } \alpha_s^i > 0$$

where some arbitrary rule is often used to resolve ties. This will be discussed further in the next section. If all $\alpha_s^i \leq 0$, then a class of solutions exists

(11) $$\sum_h a^i_{\phi(h)} (\beta^h - \theta\alpha_s^h) + \theta a_s^i = b^i \qquad (\theta > 0)$$

with the values

(12) $$\begin{aligned} \sum_h a^o_{\phi(h)} (\beta^h - \theta\alpha_s^h) + \theta a_s^o &= z - \theta \sum_{h,i} a^o_{\phi(h)} \pi_i^h a_s^i + \theta a_s^o \\ &= z + \theta \left(\sum_i \pi_i a_s^i + a_s^o\right) = z + \theta d_s < z \end{aligned}$$

and as θ approaches $+\infty$, this quantity approaches $-\infty$.

If some $\alpha_s^i > 0$, then the value of θ is bounded by θ^r which is such as to make

$$x^{\phi(r)} = \beta^r - \theta^r \alpha_s^r = 0.$$

The new values of the other basic variables, call them $\overset{*}{\beta}{}^h$, will be

(13) $$\begin{aligned} x^s &= \overset{*}{\beta}{}^r = \theta^r \\ x^{\phi(h)} &= \overset{*}{\beta}{}^h = \beta^h - \theta^r \alpha_s^h \geq 0 \quad (h \neq r) . \end{aligned}$$

We must now remove $a^i_{\phi(r)}$ from the basis and replace it with a_s^i. This is formally accomplished merely by making $\overset{*}{\phi}(r) = s$ but of course we must change the inverse. Since

$$\sum_h a^1_{\phi(h)} \alpha^h_s = a^1_s = a^1_{\bar{\phi}(r)}$$

clearly

$$(14) \qquad \sum_{h \neq r} a^1_{\phi(h)} \left[\frac{-\alpha^h_s}{\alpha^r_s} \right] + a^1_s \left[\frac{1}{\alpha^r_s} \right] = a^1_{\phi(r)} \; .$$

Now since $\bar{\phi}(h) = \phi(h)$ for $h \neq r$, and since

$$(15) \qquad \sum_{h \neq r} a^1_{\phi(h)} \pi^h_j + a^1_{\phi(r)} \pi^r_j = \delta^1_j$$

substituting (14) into (15) gives

$$\sum_{h \neq r} a^1_{\bar{\phi}(h)} \left[\pi^h_j - \pi^r_j \frac{\alpha^h_s}{\alpha^r_s} \right] + a^1_{\bar{\phi}(r)} \left[\pi^r_j \frac{1}{\alpha^r_s} \right] = \delta^1_j .$$

Thus

$$(16) \qquad \begin{aligned} \bar{\pi}^r_j &= \pi^r_j \frac{1}{\alpha^r_s} && (j = 1, \ldots, m) \\ \bar{\pi}^h_j &= \pi^h_j - \pi^r_j \frac{\alpha^h_s}{\alpha^r_s} && (h \neq r) \; . \end{aligned}$$

This, of course, is the same kind of E.R.T.s we have been making right along. Note that (13) and (16) involve the same E.R.T.s. It is also easy to show that the new pricing vector is given by

$$(17) \qquad \bar{\pi}_1 = \pi_1 - \pi^r_1 \frac{d_s}{\alpha^r_s}$$

as can be seen by considering the change

$$\overset{*}{\pi}_1 - \pi_1 = \sum_h a^o_{\phi(h)} \pi^h_1 - \sum_h a^o_{\overset{*}{\phi}(h)} \overset{*}{\pi}{}^h_1$$

$$= \sum_{h \neq r} a^o_{\phi(h)} \left[\pi^r_1 \frac{\alpha^h_s}{\alpha^r_s} \right] + \left[a^o_{\phi(r)} - a^o_s \frac{1}{\alpha^r_s} \right] \pi^r_1$$

$$= - \frac{-\pi^r_1}{\alpha^r_s} \left[-\sum_{h=1}^{m} a^o_{\phi(h)} \alpha^h_s + a^o_s \right] = - \frac{\pi^r_1}{\alpha^r_s} d_s$$

Since α^r_s is always chosen non-zero, the above E.R.T.s are all valid, hence linear independence of the rows is maintained. We have only to start with the identity matrix as the initial basis to insure the validity of all transformations. In practice, one always does start with the identity matrix either because it arises from slack vectors, because it is introduced artificially for a Phase I, or as a convenient way of inverting a given basis. The use of Phase I was discussed in the last lecture and will be taken up again in a later one. Hence we will not pursue it at this point.

DEGENERACY

We have left to the last the matter of degeneracy, which by the rule (10) can conceivably cause indefinite cycling through the same set of bases. It is logically easier to discuss at this point but it is also the least interesting part of the theory. There used to be a great hulabalu about this theoretical hole in the simplex method but nowadays people more or less ignore it and go on about their business. Unquestionably the

theory had to be completed but actual cycling in a real problem has never been reported. A few examples have been cooked up to prove it can happen. However, no proof has ever been given that the simplex method will converge in a reasonable number of iterations, cycling or not. It just does.

The proof that the simplex method will eventually reach an optimal solution depends on the value of z decreasing by a non-zero amount each iteration. Since there are only a finite number of bases and each successive one is better than the last, they must eventually terminate. The mere existence of ties in (10) does not invalidate this argument, it is only when the value of θ^r turns out to be zero that a difficulty arises. Unfortunately, the schemes that have been devised for avoiding degeneracy depend on resolving all ties, zero or not, and on maintaining feasibility at every step. We will see later that it is possible and often convenient to work with infeasible solutions and hence rigorous resolution of ties becomes impractical since it would require extremely elaborate machinery. Nevertheless, we will show how degeneracy could be avoided.

Again we make use of the identity matrix. Instead of a single column b^1, consider the more general problem restraints

$$\sum_{j=1}^{n} a_j^1 x_k^j = b_k^1 \qquad (k = 0, 1, \ldots, m)$$

where b^1 becomes b_o^1 and $b_k^1 = \delta_k^1$ $(k = 1, \ldots, m)$. (It is really only necessary that b_k^1 have linearly independent rows but in practice the identity matrix is by far the most convenient extension so we will proceed on that basis.) We will generalize

the non-negativity requirement on the variables as follows.

Consider the rows of x_k^j as lexicographically ordered vectors (L.O.V.), by which we mean that, for fixed r, and considering the whole row at once,

$$x_k^r \succ 0 \text{ if } x_o^r > 0 \text{ or if } x_o^r = x_1^r = \dots = x_{\ell-1}^r = 0.\ x_\ell^r > 0 \quad (1 \leq \ell \leq m)$$

and

$$x_k^r \succ y_k^r \text{ if } (x_k^r - y_k^r) \succ 0 \text{ in the above sense.}$$

The relation denoted by " $\succ$ " constitutes an ordering on row vectors as is easily seen from the definition.

The functional is similarly generalized to

$$z_k = \sum_{j=1}^{n} a_j^o x_k^j$$

and the elimination ratio θ^r now becomes the vector

$$(18) \qquad \theta_k^r = \min_{\alpha_s^1 > 0} \frac{\beta_k^1}{\alpha_s^1} \qquad (k = 0, 1, \dots, m)$$

the minimum being taken in the L.O.V. sense. Hence the new solution when a_s^1 is introduced into and $a_{\phi(r)}^1$ eliminated from the basis is

$$\overset{*}{\beta}{}_k^r = \frac{\beta_k^r}{\alpha_s^r} = \theta_k^r$$

$$\overset{*}{\beta}{}_k^1 = \beta_k^1 - \theta_k^r \alpha_s^1 \qquad (1 \neq r).$$

Clearly, $\overset{*}{\beta}{}_k^r \succ 0$ if $\beta_k^r \succ 0$. Similarly, for all other $1 \neq r$,

$$(19) \qquad \frac{\beta_k^1}{\alpha_s^1} \succ \frac{\beta_k^r}{\alpha_s^r} \qquad (k = 0, 1, \dots, m)$$

unless, for some $i = r'$ and constant $c = \dfrac{\alpha_s^{r'}}{\alpha_s^r} > 0$,

$$(20) \qquad \beta_k^{r'} = c\,\beta_k^r \qquad (k = 0, 1, \ldots, m).$$

But this cannot be the case, for, since the β_k^i are obtained from b_k^i by E.R.T.s and the rows of b_k^i are linearly independent, the rows of β_k^i must be linearly independent, whereas (20) would imply that with $\lambda_{r'} = 1$, $\lambda_r = -c \neq 0$,

$$\lambda_{r'}\beta_k^{r'} + \lambda_r \beta_k^r = 0 \text{ for all } k.$$

Hence (19) holds and therefore

$$\beta_k^i > \theta_k^r \alpha_s^i \qquad (i \neq r)$$

whence

$$\bar{\beta}_k^i > 0 \text{ if } \beta_k^i > 0.$$

We have thus proved that, starting with a basic solution matrix $\beta_k^i > 0$, the condition will be maintained. If we start with the identity matrix as a basis, and all $b^i \geq 0$, then the initial basic solution matrix is

$$\begin{bmatrix} b^1 & 1 & 0 & 0 & . & . & . & 0 \\ b^2 & 0 & 1 & 0 & . & . & . & 0 \\ & & \vdots & & & & & \\ b^m & 0 & 0 & 0 & . & . & . & 1 \end{bmatrix} = \beta_k^i$$

so that every β_k^i is strictly (L.O.V.) positive. Therefore, at every step, the change in the functional

$$\bar{z}_k - z_k = \theta_k^r d_s$$

is strictly (L.O.V.) negative, since $d_s < 0$ by the usual choice, and the process must eventually terminate.

The computation of π_1 and π_1^h as well as the d_j is unchanged. However, by a very happy circumstance (which we cleverly arranged)

$$\beta_k^h = \pi_k^h \qquad (h,k = 1, \ldots, m)$$

so we have no additional computation except the "generalized ratios" in (18). Even then there is no extra work unless we find, for $r' \neq r$,

$$\frac{\beta_o^{r'}}{\alpha_s^{r'}} = \frac{\beta_o^r}{\alpha_s^r}$$

in which case we examine

$$(21) \qquad \frac{\beta_1^{r'}}{\alpha_s^{r'}} \text{ vs } \frac{\beta_1^r}{\alpha_s^r} \; ; \; \frac{\beta_2^{r'}}{\alpha_s^{r'}} \text{ vs } \frac{\beta_2^r}{\alpha_s^r} \; ; \; \ldots \; ; \; \frac{\beta_m^{r'}}{\alpha_s^{r'}} \text{ vs } \frac{\beta_m^r}{\alpha_s^r}$$

in order until we find a pair of unequal ratios. As we saw following (20), all pairs cannot be equal. There may, of course, be more than two equal ratios on first components in (18). We have merely illustrated the situation for a two-way tie.

It should be noted that the choice using (21) may be made on negative ratios. In this case, the min in (18) must be understood in an algebraic sense on the "tie-breaker" element although there will, of course, be a leading positive element. If $\beta_o^r = 0$ on the current iteration, then there must have been a tie on the last iteration which was resolved to make the first

non-zero element positive, excepting of course on the first iteration which is guaranteed O.K. by the initial solution.